

Introduction to Data Management Planning

Joy Davidson

Digital Curation Centre

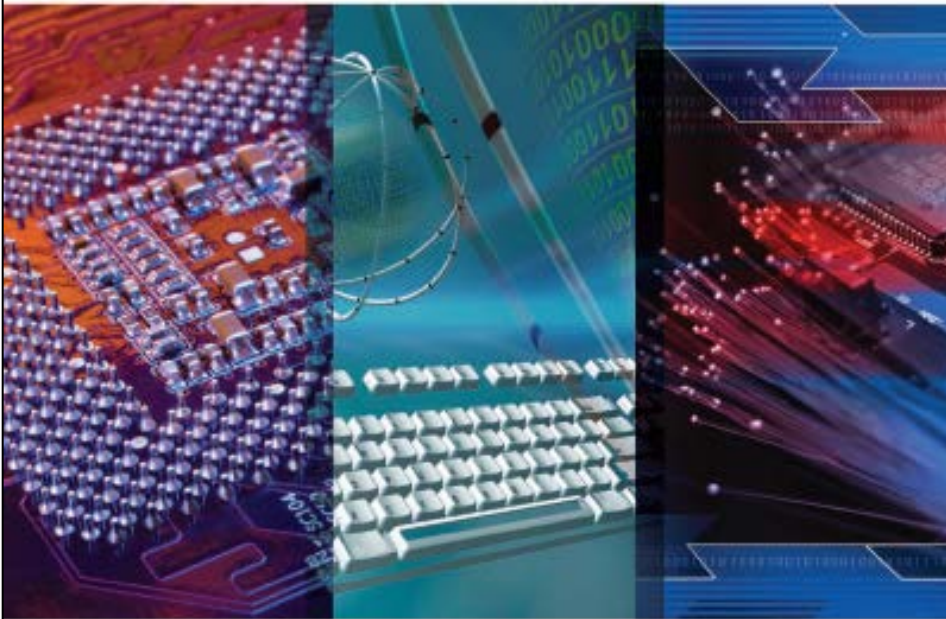
Winning Horizon2020 with Open Science.

Universidad Carlos III de Madrid. May 13, 2015



RCUK Review of e-Science 2009

BUILDING A UK FOUNDATION FOR THE TRANSFORMATIVE
ENHANCEMENT OF RESEARCH AND INNOVATION



*“Data sets are
becoming the
new instruments
of science”*

Dan Atkins, University of Michigan

Funders' expectations of public access

“Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.”

RCUK Common Principles on Data Policy

<http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

EU Policies



Guidelines on Data Management in Horizon 2020

Version 1.0
11 December 2013

Ultimately funders expect:

- **timely release of data**
 - once patents are filed or on (acceptance for) publication
- **open data sharing**
 - minimal or no restrictions if possible
- **preservation of data**
 - typically 5-10+ years if of long-term value



See the RCUK Common Principles on Data Policy:
www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

Data Sharing

What is data sharing?

“... the practice of making data used for scholarly research available to others.” [Wikipedia]

Who's involved?

- the data sharer
- the data repository
- the secondary data user
- support staff
- research participants
- commercial partners

Reasons to share data

BENEFITS

- Avoid duplication
- Scientific integrity
- More collaboration
- Better research
- Increased citation



9-30% increase shown in study

(Piwowar H. and Vision T.J 2013 ,
<https://peerj.com/preprints/1.pdf>)

DRIVERS

- Public expectations
- Government agenda
- RCUK Data Policy
www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
- UKRIO Code of Practice for Research
www.ukrio.org/what-we-do/code-of-practice-for-research/

Managing restrictions on sharing

Ethics

Balance data protection with data sharing

- Informed consent – cover current **and** future use
- Confidentiality – is anonymisation appropriate?
- Access control – who, what, when?

IPR

- Clarify copyright before research starts
- Consider licensing options e.g. Creative Commons

How to share research data

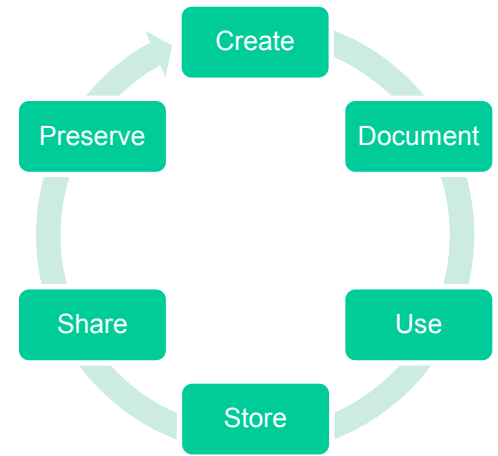
- Use appropriate repositories and data catalogues
 - <http://databib.org>
 - <http://www.re3data.org/>
 - Jisc/DCC research data registry (coming soon!)
- License the data so it is clear how it can be reused
 - www.dcc.ac.uk/resources/how-guides/license-research-data
- Make sure it's clear how to cite the data
 - <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>
- Consider publishing a data paper based on your DMP
<http://metajnl.com/>

Planning to Share

What is a DMP?

A short plan that outlines

- what data you will create and how
- how you will manage it
(storage, back-up, access...)
- plans for data sharing and preservation



DMPs are often submitted as part of grant applications, but are useful whenever you're creating data.

Why develop a DMP?

- to help you manage your data
- to provide guidelines for everyone to work to
- to anticipate and avoid problems e.g. data loss
- to avoid duplication, data loss & security breaches
- to comply with funders requirements...

Research funders have DMP requirements

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	○
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	○	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	○
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

They typically want a short (c.1-2pp) statement covering:

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation





CRediT

An open standard for expressing roles intrinsic to research

<http://credit.casrai.org/>

#1 conceptualization

Ideas; formulation or evolution of overarching research goals and aims.

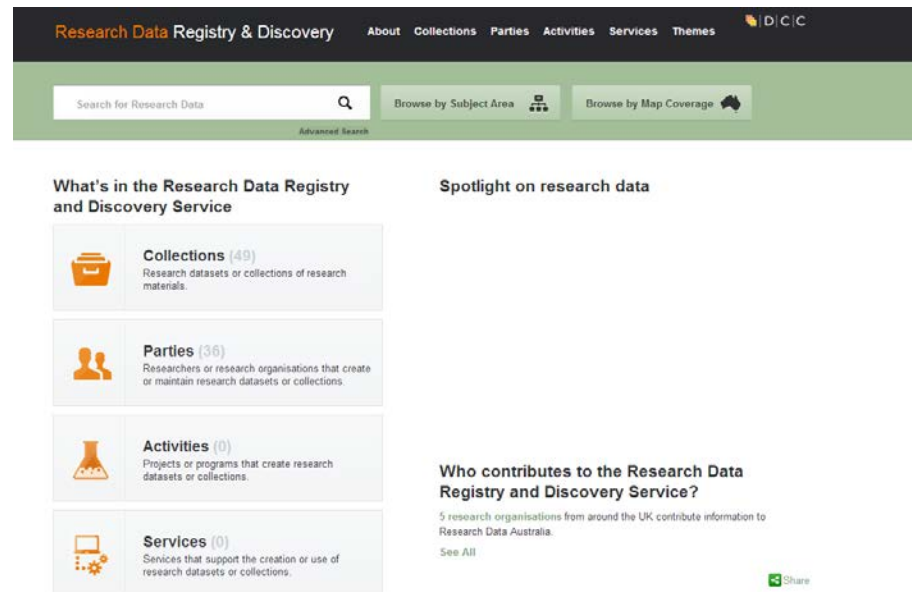


<https://dmponline.dcc.ac.uk/>

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://www.re3data.org/>

Literature and data review, data management planning



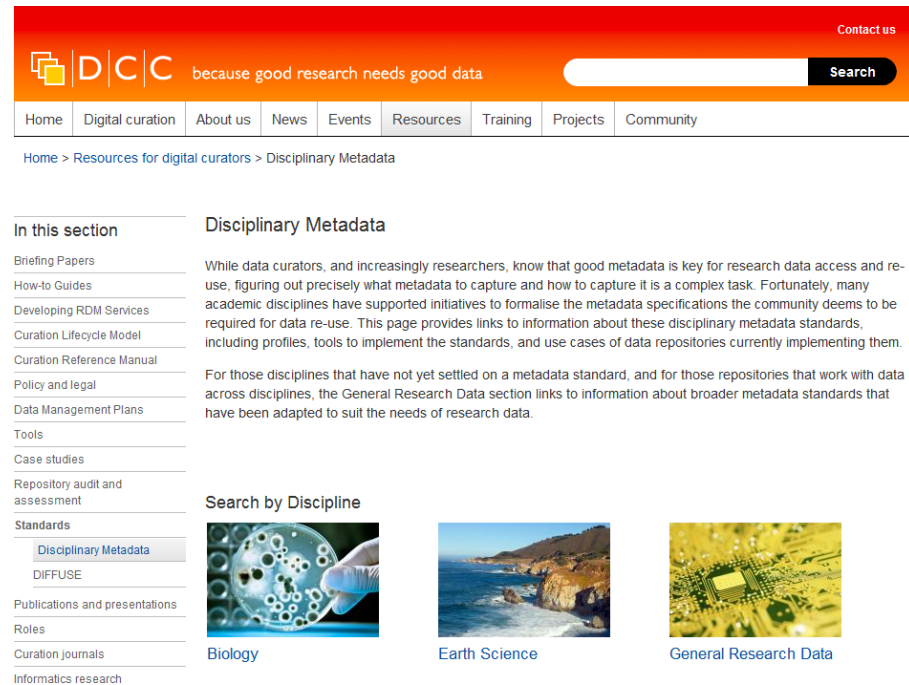
<http://rdrds.cloudapp.net/registry/>

#2 methodology

Development or design of methodology; creation of models.

Standards, formats, consent

Informed consent



Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Search by Discipline

- Biology
- Earth Science
- General Research Data

<http://www.dcc.ac.uk/resources/metadata-standards>

#4 validation

Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.

Selecting and appraising data to
be retained



#5 formal analysis

Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.

Describing techniques in data catalogue records



<http://www.software.ac.uk/>



DATA CARPENTRY

MAKING DATA SCIENCE MORE EFFICIENT

<http://datacarpentry.org/>

#6 investigation

Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

The image shows a four-panel interface for selecting a Creative Commons license. The panels are arranged in a 2x2 grid, connected by arrows indicating a sequential flow from top-left to top-right, then down to bottom-right, and finally down to bottom-left.

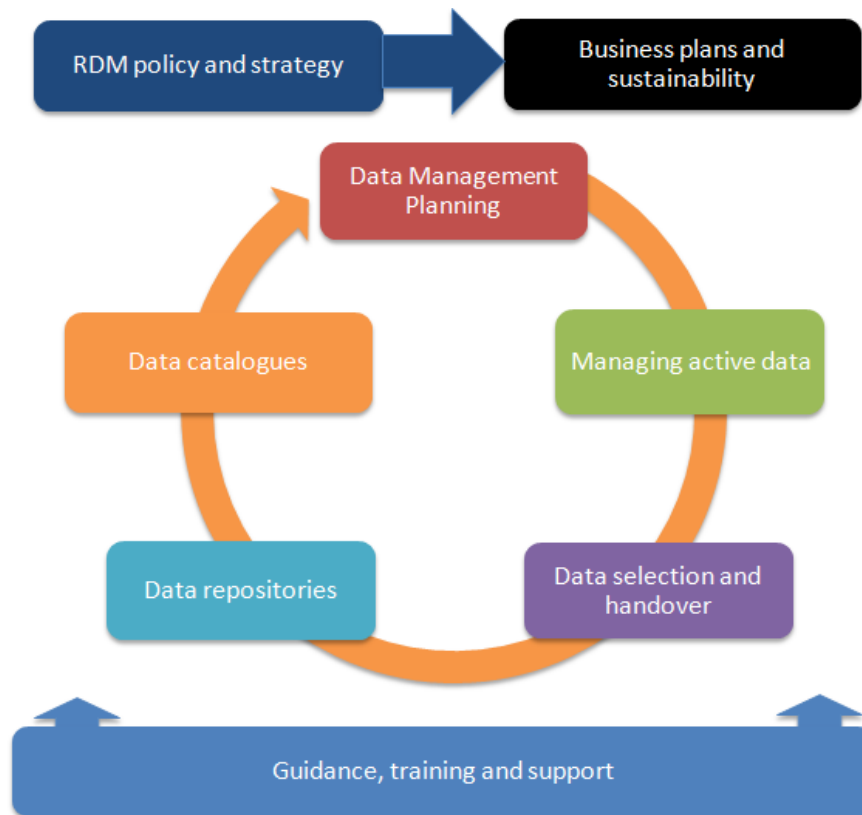
- License Features:** Contains two sections with radio button options. The first section, "Allow adaptations of your work to be shared?", has options for "Yes" (selected), "No", and "Yes, as long as others share alike". The second section, "Allow commercial uses of your work?", has options for "Yes" (selected) and "No".
- Selected License:** Displays the chosen license as "Attribution 4.0 International". It features the CC logo, the person icon, and a green circular badge that says "Free Culture APPROVED FOR Works". Below the icons, it states "This is a Free Culture License!".
- Help others attribute you!:** This panel is optional and contains several input fields: "Title of work", "Attribute work to name", "Attribute work to URL", "Source work URL", and "More permissions URL". It also has a dropdown menu for "Format of work" (set to "Other / Multiple formats") and a dropdown for "License mark" (set to "HTML+RDFa").
- Have a web page?:** Shows the license icon (CC BY) and states "This work is licensed under a Creative Commons Attribution 4.0 International License." It includes a section "Copy this code to let your visitors know!" with a text area containing HTML code for embedding the license on a website. At the bottom, there are radio buttons for "Normal Icon" (selected) and "Compact Icon".

Data collection; reuse of third party data; searching for existing datasets for reuse

<http://creativecommons.org/choose/>

#8 data curation

Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.

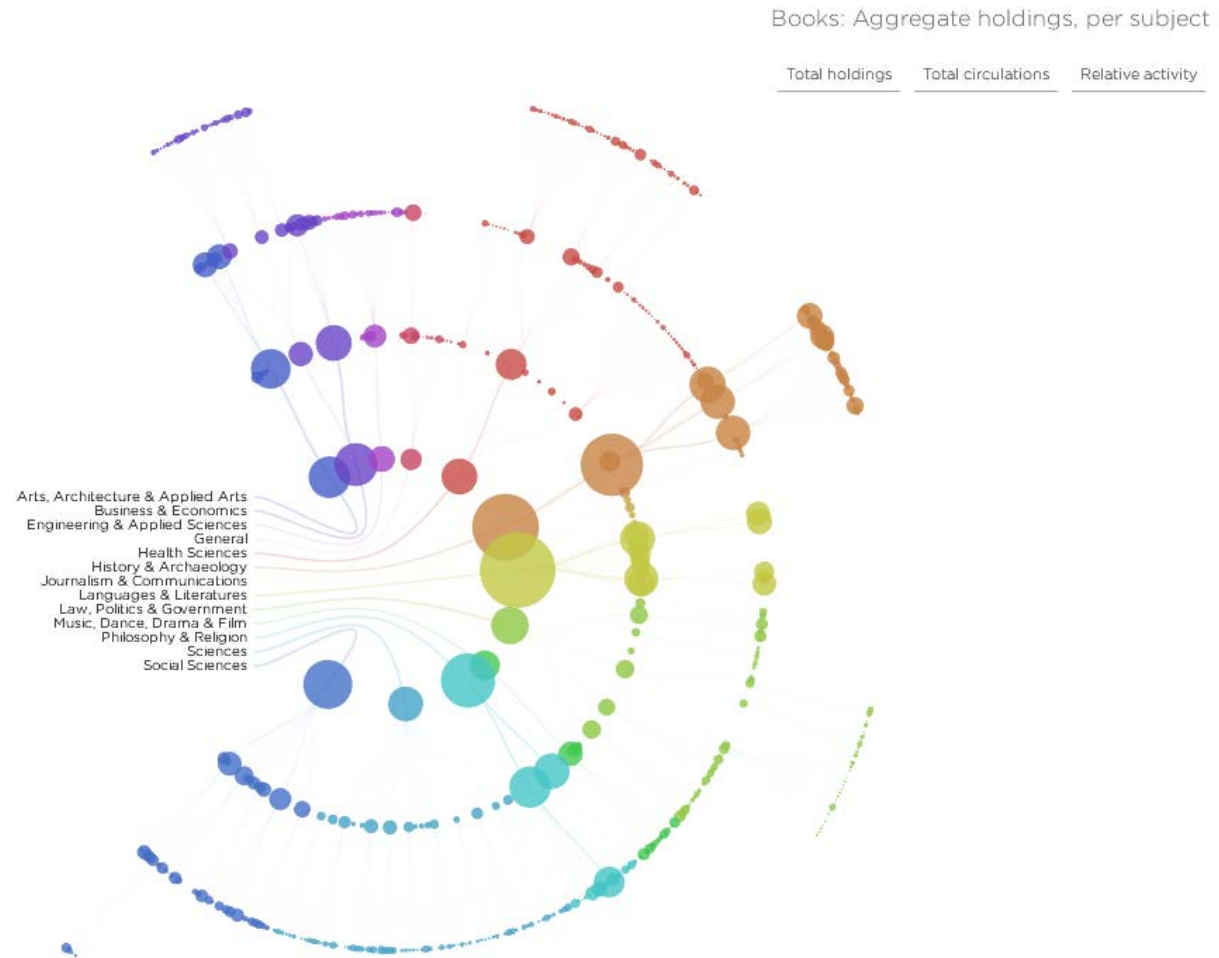


Determine curation requirements

#11 visualization

Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.

Data models and visualisations



http://spatialinformationdesignlab.org/project_sites/library/catalog.html

#12 supervision

Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.



<https://www.fosteropenscience.eu/>

Training, supervising and
advising other researchers



<http://www.researchinfonet.org/infolit/ridls/>

#13 project administration

Management and coordination responsibility for the research activity planning and execution.

Understanding institutional
and funding body
requirements



<http://gtr.rcuk.ac.uk/>

<https://www.researchfish.com/>

UCT POLICY for RESPONSIBLE CONDUCT OF RESEARCH

[Last revised August 2012]

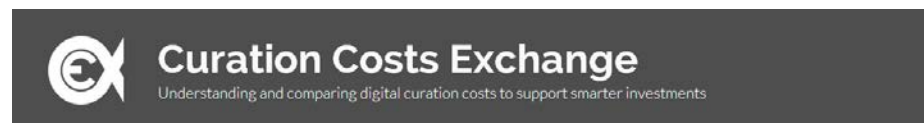
This document must be read with the UCT Authorship Practices Policy, the UCT Research Ethics Code for Research Involving Human Participants, the UCT Research Ethics Code for Use of Animals in Research and Teaching, the [DRAFT] UCT Policy for Conflicts of Interest and of Commitment in Teaching and Research, the [DRAFT] UCT Policy and Procedures for Breach of Research Ethics Codes and Allegations of Research Misconduct in Research, the [DRAFT REVISED] Policy for Avoiding Plagiarism, the [DRAFT] Whistle-blowing for Academic Misconduct Policy.

<https://www.uct.ac.za/downloads/uct.ac.za/about/policies/UCTresearchconductpolicy.pdf>

#14 funding acquisition

Acquisition of the financial support for the project leading to this publication.

Aware of funding body mandates relating to RDM and data sharing and related costs



Home Get started Compare costs Understand your costs Read more Discuss and share Related services Sign Up Sign In

All about the costs of curation

Understand what you and others are and should be spending in digital curation

Compare costs
Understand what you and others are and should be spending in digital curation



Understand costs

Understand how to assess your curation costs and how to make use of cost models to help you invest

<http://www.curationexchange.org/>

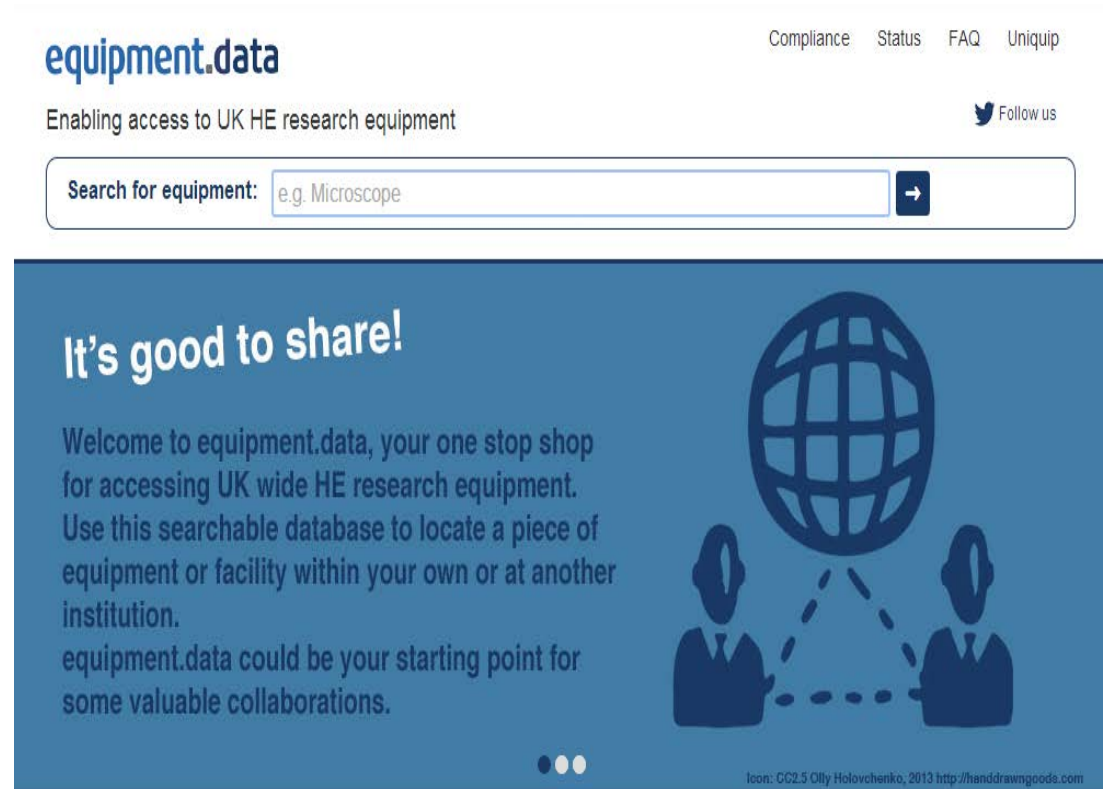
The screenshot shows the REF2014 Research Excellence Framework website. The header includes the REF2014 logo and a navigation menu with links like Publications, Submissions, Expert panels, Research users, Equality & diversity, Background, Timetable, FAQs, and Contact. The main content area is titled "Research Excellence Framework" and contains text about the new system for assessing research quality, replacing the Research Assessment Exercise (RAE). It mentions that the REF will be undertaken by the four UK higher education funding bodies and that the primary purpose is to produce assessment outcomes for each submission. A sidebar on the right titled "REF 2014 Latest" provides updates, including the publication of outcomes on 18 December 2014 and the submission deadline of 29 Nov 2013.

<http://www.ref.ac.uk/>

#7 resources

Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.

Documenting research
methods and processes



<http://equipment.data.ac.uk/>

#9-10 writing – original draft; review and editing,

Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).

Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.



ORCID
Connecting Research
and Researchers

FOR RESEARCHERS FOR ORGANIZATIONS ABOUT HELP SIGN IN

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. [Find out more.](#)

- 1 REGISTER** Get your unique ORCID identifier. Register now! Registration takes 30 seconds.
- 2 ADD YOUR INFO** Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or iD).
- 3 USE YOUR ORCID ID** Include your ORCID identifier on your Webpage, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work.

iD

LATEST NEWS
Wed 07/04/2014
A recipe for trust
Wed 06/11/2014

DC¹
Data Citation Principles

<https://www.force11.org/datacitation>

Publishing



UP metaJournals
Data

Home About

DATA
Data papers highlight openly archived data with high reuse potential, and provide recognition for the producers of the data.

SOFTWARE
Software papers help you to locate openly archived, reusable code relevant to your research, and provide a mechanism for citing its use.

RESEARCH
Research reports provide concise summaries of key developments in a field that the community do not otherwise have access to.

Journal of Open Archaeology Data | Journal of Open Psychology Data | Open Health Data | Journal of Open Research Software

Copyright Ubiquity Press | CC All content is licensed under a Creative Commons License



DataCite
Helping you to find,
access, and reuse data
<https://www.datacite.org/>

<http://www.metajnl.com/>

Guidance and tools



INICIO CREAR UN PGD PGDonline DOCUMENTOS FAQs ASesoramiento

PA GO DA - Plan de Gestión de Datos

Crear su Plan de Gestión de Datos

El Plan de Gestión de Datos lo solicita un agente financiador como parte de las condiciones del contrato de subvención para un proyecto científico.

El Programa Horizonte 2020 requiere que los proyectos que formen parte del Piloto de Datos de Investigación en Abierto entreguen un Plan de Gestión de Datos completo durante los 6 primeros meses del proyecto.

Los planes de gestión de datos son una parte integral de las solicitudes de subvenciones - no pueden ser una idea de último momento; los revisores buscarán evidencia de que la gestión de datos está incluida en su propuesta, y que forma parte integral de su proceso de investigación. En el artículo 29.3 del [H2020 Model Grant Agreement: Multi-beneficiary General MGA: December 2013](#) se establecen las obligaciones de los participantes en el Piloto de Datos de Investigación en Abierto en lo que respecta a la gestión de los datos.

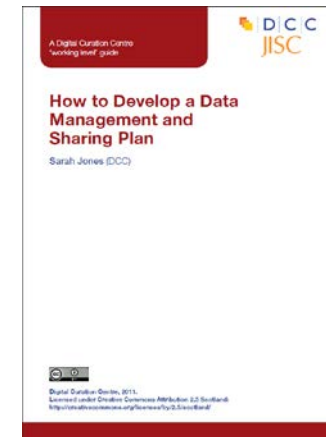
El documento [Directrices sobre la Gestión de los Datos en Horizonte 2020](#) se dirige a los solicitantes y beneficiarios de los proyectos en el Marco del Piloto de Datos de Investigación en Abierto y su objetivo es proporcionar indicaciones sobre cómo pueden cumplir con sus responsabilidades con respecto a la calidad de los datos de investigación, su intercambio y su seguridad.

Para saber más sobre el Piloto de Datos de Investigación en Abierto, consulte nuestras [Faq's](#)
Para saber más sobre el PGD como parte del Piloto de Datos de Investigación en Abierto, consulte nuestras [Faq's](#)
Para saber más sobre el Horizonte 2020, consulte [estos documentos](#)

Recomendaciones para la preparación de su Plan de Gestión de Datos (PGD):

Antes de empezar a crear su Plan de Gestión de Datos es muy conveniente que previamente considere una serie de cuestiones relacionadas con la gestión de los datos en su proyecto.

Un punto de partida útil es [la lista de control DCC \(Digital Curation Centre\) para un Plan de Gestión de Datos](#) (en inglés) que podemos encontrar dentro de la selección de [recursos del DCC](#) (en inglés). Presenta los principales temas y preguntas que los investigadores pueden querer cubrir a la hora de escribir un Plan de



<http://pgd.consorciomadrono.es/>
www.dcc.ac.uk/resources/data-management-plans

Tips for writing DMPs

- Start early!
- Do not to write the plan in isolation – seek advice from colleagues (ethics, IT, library)
- Be realistic - base plans on available skills & support
- The plan will - **and should** - change over time!

Thanks – any questions?

DMP guidance, tools and resources:

www.dcc.ac.uk/resources/data-management-plans

Follow us on twitter:

@digitalcuration and #DMPonline